The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

W. Bragg
Nobel Prize in physics 1915

# 2. Background and related work

## 2.1 Augmented reality displays

The augmented reality prototype built by Sutherland in 1968 used a head-mounted display as illustrated in Figure 3. These head-mounted displays required expensive equipment, limiting augmented reality to labs and research institutions. Indeed this is where augmented reality stayed until the rise of mobile computing platforms which researchers equipped with external sensors for tracking. Recent mobile computers such as smart phones are already equipped with sensors, namely GPS, digital compasses and cameras, enabling the realization of augmented reality.

Generally, AR displays can be split into head-mounted displays (HMD), handheld displays and projection displays, the latter being stationary and potentially able to accommodate multiple users.

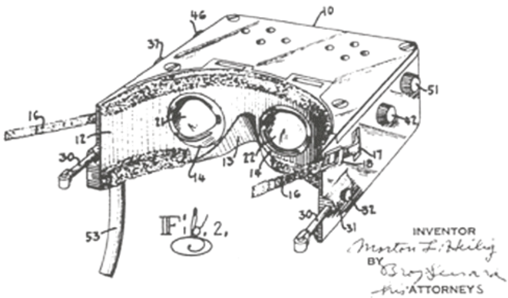


**Figure 3: The first head-mounted display by Ivan Sutherland.**

G. Schall, *Mobile Augmented Reality for Human Scale Interaction with Geospatial Models*,
Mobile Computing, DOI 10.1007/978-3-658-00197-1_2,

Also, for image generation and merging with the real world, two approaches can be distinguished as described by (Schmalstieg and Reitmayr 2006): optical see-through systems, which allow the user to see through the display onto the real world, and video see-through systems, which use video cameras to capture an image of the real world and provide the user with an augmented video image of the environment. As a result, five major classes of AR can be distinguished by their display type and their merging approach: optical see-through HMD AR, video see-through HMD AR, handheld display AR, projection-based AR with video augmentation and projection-based AR with physical surface augmentation.

Projection-based AR with video augmentation uses video projectors to display the image of an external video camera augmented with computer graphics on the screen whereas projection-based AR with physical surface augmentation projects light onto arbitrarily shaped real-world objects. It uses the real-world objects as the projection surface for the virtual environments. Ordinary surfaces have varying reflectance, color and form. Limitations of mobile devices, such as low resolution and small field of view, focus constraints, and ergonomic issues can be overcome in many cases by the utilization of projection technology. Applications that do not require mobility can benefit from efficient spatial augmentations. The focus of the work is on mobile AR systems.

## 2.2 Mobile augmented reality

A mobile AR system can present three-dimensional information superimposed on a roaming user's view of a task location. A decade ago, HMDs were widely built and employed by research groups. More recently, the focus has shifted towards smaller handheld devices, as depicted in Figure 4 showing the evolution of mobile AR systems. Head-mounted displays are usually worn by the user and provide two image-generating devices, one for each eye. Optical see-through HMD AR uses a transparent HMD to blend together virtual and real content. Video see-through HMD AR uses an opaque HMD to display merged video of the virtual environment with and view from cameras on the HMD. By overlaying the video images with the rendered content before displaying both to the user, virtual objects can appear opaque and occlude the real objects behind them.

Early work on mobile AR, such as the Touring Machine from (Feiner, MacIntyre, Höllerer, & Webster, 1997) used backpacks with laptop computers and HMDs (see Figure 5). (Höllerer, Feiner, Terauchi, Rashid, & Hallaway, 1999) built a series of mobile AR systems (MARS) prototypes, starting with extensions to the Touring Machine.

**Figure 4: Evolution of mobile AR systems. Hardware setups range from backpack systems to handheld computers.**

Similar augmented reality prototypes have been built by Piekarsky and Thomas in form of the Tinmith system (Piekarski & Thomas, 2001). The Tinmith-Metro application is the main application that demonstrates the capture and creation of 3D geometry outdoors in real-time, leveraging the user's physical presence in the world. Furthermore, (Reitmayr & Schmalstieg, 2004) have shown a collaborative augmented reality application for outdoor navigation and information browsing. However these systems are rather cumbersome for mobile applications deployed over longer working periods. Research focused on smaller, more convenient prototypes.

With the advent of handheld devices featuring cameras the video see-through metaphor has been widely adopted for AR systems providing augmented or "X-ray vision" views to the user. Consequently, handheld AR displays also use the video-see-through approach (Schmalstieg & Reitmayr, 2006). However they can be built from tablet PCs, Ultra Mobile PCs, or even mobile phones and devices which are highly available, and have good technical and ergonomic acceptance.


**Figure 5: Mobile AR systems. (Left) Situated Documentaries. User with MARS prototype based on backpack including differential GPS. (Right) Tinmith prototype from University of South Australia.**

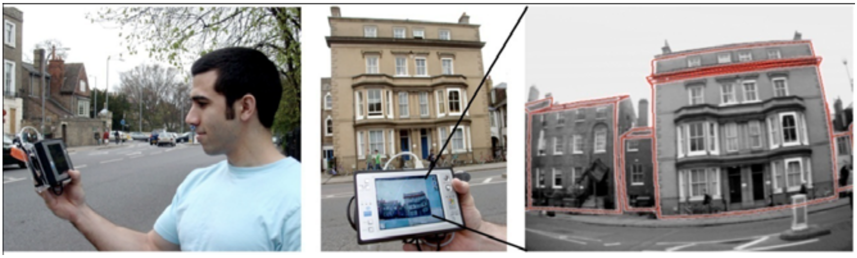

**Figure 6: Going out. (Left) A user operating a handheld augmented reality unit tracked in an urban environment. (Middle) Live shot showing the unit tracking a building. (Right) Screenshot from a pose close to the left images with overlaid building outline.**

Recently, handheld AR displays became popular and can be potentially used in ubiquitous computing. This alternative and more ergonomic approach based on a handheld computer was originally conceived by (Fitzmaurice & Buxton, 1994), and later refined into a see-through AR device by (Rekimoto, 2001). UMPCs are basically small mobile PCs running standard operation systems. This has started a strong trend towards handheld AR (Wagner, Pintaric, Ledermann, & Schmalstieg, 2005). For example, Kruijff and Veas designed a two-handed shell around an Ultra Mobile PC (Kruijff & Veas, 2007).

Moreover, Reitmayr and Drummond demonstrated a vision-based tracking approach on an UMPC as depicted in Figure 6 (Reitmayr & Drummond, 2006). Today, already smart phones are fully featured high-end cell phones with GPS, camera, inertial sensors and a GHz processor, so that applications for data processing and connectivity can be installed on them. These sensors reflect the state of the sensors used for early backpack AR prototypes. As the processing capability of smart phones is improving, this enables a new class of AR applications which use the camera also for vision-based tracking. Notable examples are from (Wagner, Reitmayr, Mulloni, Drummond, & Schmalstieg, 2008) utilizing them as final mobile AR displays (see Figure 7). In 2009 a promising approach was implemented within the Wikitude project (Breuss-Schneeweis, 2009), basically implementing a mobile AR travel guide with AR functionality based on user-generated Web2.0 Wikipedia or Panoramio content. The user sees an annotated landscape, mountain names or landmark descriptions in an augmented reality camera view. The problem with such approaches is that tracking solely relies on GPS and magnetometer which is leading to a poor registration. Therefore, latest research on smart phones focuses on vision-based tracking of natural features to overcome these drawbacks. The approaches should allow tracking the user in unprepared and unconstrained environments. Also Rohs used smart phones for markerless tracking of magic lenses on paper maps in real-time (Rohs, Schöning, Krueuger, & Hecht, 2007).

Consider that mobile AR can be realized on a variety of hardware platforms depending on the user group, requirements and the specific tasks. Among the recurring themes of AR research are world-registered (augmentable) annotations. Mobile AR is specifically suited for mobile spatial interaction (Fröhlich, 2009). Experience showed that the mobile user is very interested in interacting with the AR models. For example, (Thomas & Piekarski, 2002) experimented with spatial interaction with 3D models (see Figure 5 (right)). Lately, (Wither, DiVerdi, & Höllerer, 2009) investigated annotations for augmented reality. Typically, mobile spatial interaction is performed on superimposed geospatial 3D models or new annotations are spatially fixed to the real-world view. (Paelke & Brenner, 2007) investigate interaction with spatial data considering which tasks are relevant and consider the scope of the interaction. Interaction tasks include identification of objects, information about objects, localization of objects, user guidance, navigation, spatial selection, spatial positioning, and data collection.

Such procedures enable users with the capability of for example on-site documentation, interactive placement or correction of information. For example, in industrial settings this would be useful for field workers of utility companies aiming to locate particular items of the underground infrastructure. More generally, this procedure is also useful for city tourists using their smart phone for leaving annotations in the space. Next, the author examines what kind of data and data sources can be utilized for building geospatial 3D models that can be used for AR.

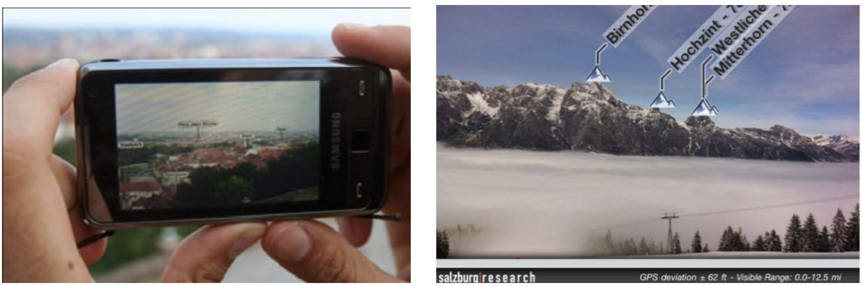


**Figure 7: A user operating a smart phone using an AR application for visualizing labels registered on the environment.**

## 2.3 Geospatial models

Mobile augmented reality requires geospatial data to present world-registered overlays. AR has a strong demand for a content generation pipeline. Currently, the process of generating 3D models for AR is not fully investigated. A leak of models for AR can be a bottle-neck for the future growth of AR applications. It seems reasonable to exploit already existing data stored in databases. Furthermore, large productive geospatial databases are the result of hundreds of person years of surveying effort. For example, a procedure of turning raw geospatial data, which are mostly 2D, into 3D models suitable for standard rendering engines could help providing manifold models for AR.

Geospatial data – also known as geographic data – refers to a particular kind of data, which is spatially referenced to the surface of the earth. The data is typically organized in geodatabases, which in turn are implemented and managed using geospatially enabled database management systems (GeoDBMS). (Schmalstieg et al., 2007) proposed a pipeline for managing AR models along the lines of a conventional information processing pipeline, which has as its main stages acquisition, storage, delivery, and use of the data. This organization separates creation and use of AR data into distinct phases. The long-term goal of mobile AR is to let users move unconstrained throughout a wide-area, and to continuously provide assistance for a wide variety of tasks. This requires coverage of the whole area and all the possible contained tasks in the underlying AR model. Scaling AR models to such wide-area-modeling coverage is only practical by leveraging legacy databases, such as existing digital maps. Manual methods for the creation of 3D models for AR are typically time consuming (see Figure 8).

The most common way to interact with geospatial data held in geodatabases is by means of a geographic information system (GIS). According to (Bruenig & Zlatanova, 2006) a GIS is a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world. Sophisticated tools for spatial analysis permit to generate information relevant to decision making from the data held in underlying geodatabases. GIS can also connect to data sources other than geodatabases such as satellite imagery provided in a specific file format or Web services delivering imagery or collections of features. GIS play a major role in the context of spatial asset management (utilities and telecommunication), mapping and cadastral surveying, navigation and location- based services, planning and spatial business analysis. GIS have been available since the late 1970s by then running as monolithic stand-alone systems. In the 1990s GIS shifted towards desktop-based but still stand-alone applications. Recent developments show an increasing integration of GIS into enterprise-wide solutions where GIS communicates directly with other systems by means of Web services. In resent years a trend towards mobile GIS and 3D GIS is observable. Exploiting data stored in GIS allows for rapid and on the fly generation of three-dimensional representations of the data. Moreover, data sources such as CAD construction drawings present a huge reservoir for semantic geospatial models that can be extracted and applied in AR applications.

The vision can also be inferred from the trends in GIS research mentioned by geoscientists (Huisman & Forer, 2008):

– The increasing availability and use of shared derivative data artifacts;
– The increasing demand for temporal and dynamic functionality in geo-information;
– The increasing seeking for representations of objects true to their nature.

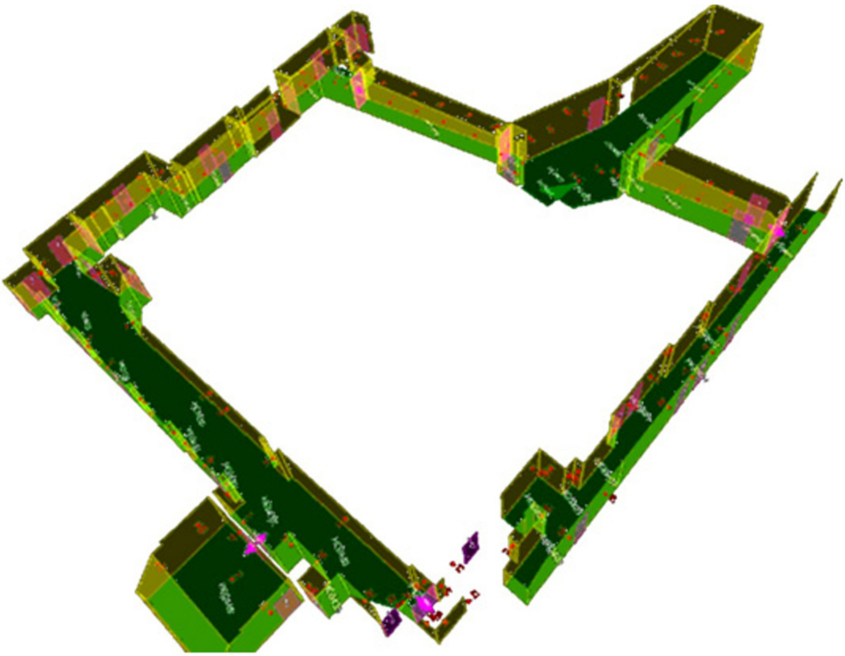

**Figure 8: Manually generated 3D model of a building floor. The model includes a corridor and few dozens rooms.**

This raises various issues with regard to the deployment of geospatial data. Specifically, the last trend highlights the importance of AR for providing realistic 3D visualizations for mobile GIS applications. Three-dimensional representation and visualization of geospatial environments are employed in an increasing number of applications, such as urban planning, urban marketing and emergency tasks. Existing urban 3D models can differ for example in data formats, level-of-detail (LoD) or type of data they are based on. Figure 9 shows a detailed 3D model of a city. Typically for outdoor visualizations in AR very simple models, such as building wireframes, are used.

Users often expect reliable data representations, so strict dependence on real-world measurements is necessary. Consequently, data formats based on standard GML ("Geography Markup Language | OGC®", 2010) are suitable. There are derivatives of GML, such as CityGML (Kolbe, Gröger, & Pluemer, 2008), which is a specialization of the GML language for 3D visualization 3D city models requiring a special browser. Instead, also a standard scene-graph structure can be used which enables to preserve the semantic data from the geo-database in the resulting 3D models. This has the advantage that semantic information can be used to change the appearance of the 3D model in real-time. (Mendez et al., 2008) describes such visualization techniques in more detail. There has been other work on forwarding database information to scene-graphs with a database, for example X-VRML (Walczak & Cellary, 2003), but these types of approaches generally do not involve on the fly procedural modeling. Storing the model data in a geospatial database provides the user with all the advantages of geo-databases, such as data access control, data loss prevention etc. (Bruenig & Zlatanova, 2006). Furthermore, the pipeline approach creates considerable added value from an economic point of view since a geospatial database can be used by many visualization applications (Schmalstieg et al., 2007). In addition, redundancy and inconsistency among spatially overlapping models are eliminated since all models are generated with reference to the most up-to-date data.

The work in (Roberts et al., 2002) seems to be the only AR application that is explicitly concerned with exploiting GIS data of underground infrastructure. (Paelke & Brenner, 2007) presented an AR device for interactive on-site visualization of geospatial models. Handheld devices exist that have been used in the exploration of GIS data. These include ARVino, exploring viticulture data (King, Piekarski, & Thomas, 2005), and simple landscape visualization system (Priestnall & Polmear, 2006). For most users the pure visualization of geospatial 3D models can be seen as the basic use case.

But even more useful is the ability to interact with the geospatial model and annotate a model.

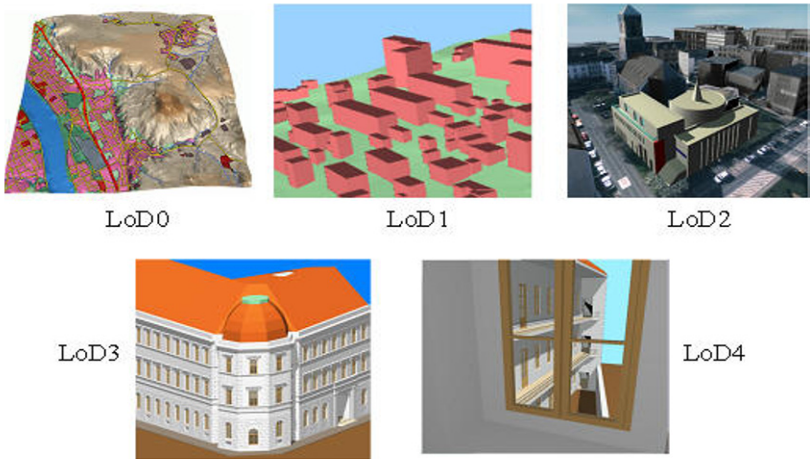


**Figure 9: 3D city model and building models in various detail based on CityGML.**

Note that AR models need to fulfill requirements for both visualization and tracking, thus including visual and non-visual information. Widespread future adoption of augmented reality technology will rely on a broadly accessible standard for authoring and distributing content with, at a minimum, the flexibility and interactivity provided by current Web authoring technologies.

For example, (Hill, MacIntyre, & Gandy, 2010) introduced KHARMA, an open architecture based on KML for geospatial marker and relative referencing combined with standard browser supported HTML5 and JavaScript technologies for content development and delivery. However all the approaches mentioned above assume that the data for generating 3D models are available at the required level of detail and accuracy.

Another important question is: "How can a user get access to geospatial models in a ubiquitous environment"? The real-time delivery of maps over the Internet to mobile users is still in its infancy. Increasing interactivity requires that the Web-based infrastructures enable the delivery of both 2D and 3D geospatial data to the mobile user. In this context, multiple representations of geospatial objects linked the ones with the others are desired to allow navigation at different levels of detail, representation or scales. Moreover, the representations of digitalized or independently captured data need to be consistent. Additionally, online processes, also called Web services, need to be available to enable the real-time delivery, analysis, modification, derivation and interaction with the different levels of scale and detail of the geospatial data. Current geospatial Web services are very often limited to those specified by the Open Geospatial Consortium (OGC) (Welcome to the OGC Website | OGC, 2010) and standardized by ISO, namely the Web Map Service (de La Beaujardiere, 2004) (service for the online delivery of 2D maps), Web Feature Service (Vretanos, 2002) and Web Coverage Service (Whiteside & J. Evans, 2006) (services for the online delivery of respectively geospatial vector and raster data). However according to (Badard, 2006), if these services constitute the essential building blocks for the design of distributed and interoperable infrastructures for the delivery and access to geospatial data, no processing, such as online analysis or creation of new information is possible. To overcome these shortcomings various geospatial service oriented architectures were investigated by (Badard, 2006).

On demand Web services for map delivery or services such as Google Earth provide maps of cities to mobile users. In addition, Internet GIS applications in planning and resource management have become more widespread in recent years. This allows for nomadic access of GIS services anyplace and anytime via the internet by using a simple web browser. Already a growing number of companies from various sectors, such as the utility or transportation sector, rely on Web applications to provide their data to construction companies or customers. In this context, Internet GIS enables mobile field workers to consult the mobile GIS at the inspection site. For example, the Austrian utility company Innsbrucker Kommunalbetriebe provides a Web interface where registered users can mark the target area on the map by drawing a polygon around the area of which they want to extract information about buried assets, such as sewer pipes, electricity or water lines. AR as a novel user interface promises to go one step further and allows viewing geospatial content in relation to the real world on-site by overlaying the virtual information over the video footage. One essential question is how to generate such geospatial content or models. Here, the Web can serve as an important pool of geospatial data.

## 2.4 Pose tracking

Any augmented reality application relies on some kind of tracking the user's or display's pose in order to register its content in respect to the real world. This means, determining position and orientation of an object is often referred to as six-degree-of-freedom (6DoF) tracking, for the six parameters sensed: position in x, y, and z, and orientation in yaw, pitch, and roll angles. 6DoF pose tracking must run in real-time, typically requiring solutions that estimate poses in less than 50 milliseconds. Furthermore it must be robust under many conditions. In case tracking is lost, the system must be able to recover quickly. Much work in mobile AR has focused on wide-area tracking. Most commercial solutions such as optical or infrared trackers cover only a limited work area, so researchers have aimed at using for example GPS (Höllerer, Feiner, Terauchi, Rashid, & Hallaway, 1999), inertial sensors (Bachmann et al., 2002), and vision (Ribo et al., 2002) for tracking. The Bat System (Newman, Ingram, & Hopper, 2001) from AT&T allows building-wide accurate tracking of people and objects outfitted with badges that are tracked by a 3D ultrasonic location system, but at the cost of deploying a building-wide electronic infrastructure.

Marker tracking is often used in AR applications if limited computational resources do not permit robust markerless tracking. One of the first projects using camera-based 6DoF tracking of artificial 2D markers was ARToolKit (Kato, Billinghurst, Blanding, & May, 1999) which was released under the GPL license and therefore became enormously popular among AR researchers and enthusiasts alike. It pioneered the use of a square planar shape for pose estimation and an embedded 2D barcode pattern for distinguishing markers. Rekimoto's 2D Matrix Code (Rekimoto, 2002) used a similar approach. Since then, many similar square tracking libraries have emerged among which the most prominent ones are ARToolKitPlus library (Wagner & Schmalstieg, 2007)(see Figure 10 (left)).

Natural feature tracking in real-time became feasible on mobile computers since recently processing power has reached a level that allows for vision-based tracking. These approaches solve the problem of polluting the user's environment with fiducial markers. Some examples are: (Bleser, Wuest, & Strieker, 2007) uses a 3D CAD model to initialize the tracking process.

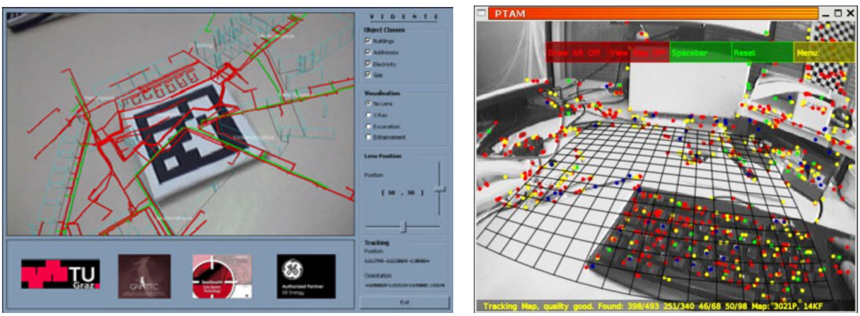


**Figure 10: Table top visualization. (Left) 3D model of underground infrastructure and wireframe buildings superimposed on a fiducial marker. (Right) PTAM (Parallel Tracking and Mapping) is a camera tracking system for augmented reality. It requires no markers, pre-made maps, known templates or inertial sen**sors.

Springer
http://www.springer.com/978-3-658-00196-4

Mobile Augmented Reality for Human Scale Interaction
with Geospatial Models
The Benefit for Industrial Applications
Schall, G.
2013, XXIII, 171 p. 83 illus., 8 illus. in color., Softcover
ISBN: 978-3-658-00196-4